Pharmacology based toxicity assessment : towards quantitative risk prediction in humans

Sahota, T.

CHAPTER 4

Application of optimal design concepts to experimental protocols for the evaluation of toxicokinetics and safety thresholds

Tarjinder Sahota, Núria Buil-Bruna, Meindert Danhof and Oscar Della Pasqua

Submitted for publication (Pharmaceutical Research)

Abstract

Purpose: In toxicology experiments measures of drug exposure are calculated using non-compartmental methods, despite evidence that population pharmacokinetic (PK) modelling can provide accurate estimates of the parameters of interest. Here we explore the utility of optimised protocol design and PK modelling on the precision of exposure measures for a variety of hypothetical compounds.

Methods: Optimal design concepts were applied to a range of hypothetical drugs with different pharmacokinetic profiles. Protocol designs were optimised both in terms of sampling schedule and number of animals per group. The precision of secondary parameters, namely AUC and C_{MAX} was used as target for optimization purposes. Adequate precision levels were defined as expected CV% < 40%. Absolute changes in expected precision of less than 10% were deemed acceptable.

Results: Independent of differences in drug disposition, our results show that the number of animals used in experimental protocols can be reduced by 2/3 with acceptable loss of precision in AUC and C_{MAX} estimates. Even though some PK parameters were found to be imprecisely estimated when drug disposition involves more than one compartment, this does not significantly affect the secondary parameters describing systemic exposure, which showed adequate precision (all CVs <36%).

Conclusions: The accuracy and precision of measures of systemic exposure such as AUC and C_{MAX} are essential to ensure appropriate interpretation of experimental findings and make inferences about safety risk in humans. However, our analysis reveals that for composite methods, which are commonly used in toxicology protocols, sample size does not determine the precision of the pharmacokinetic parameters of interest. Rather, it is the sampling scheme and dose levels which matter. In contrast to current practice, precise calculation of safety thresholds can be obtained with a considerable reduction in the number of animals used in a typical protocol.

Introduction

Despite the evidence for important limitations in the assessment of non-clinical safety and toxicology, experimental protocols and data analysis have not advanced in the same way risk management concepts have evolved over the last decade (81). Drug exposure remains a proxy for risk even when other markers of safety and toxicity might be better predictors of adverse drug reactions (5). In fact, the establishment of safe exposure levels prior to first time in human studies is still one of the most important milestones in drug development (6,7). Yet, the reliability of these estimates depends on the quality, accuracy and precision of the data obtained from preclinical toxicology experiments. Even though statistical considerations are described in current guidelines, these methodological aspects appear to remain beyond the scope of the scientific debate on the relevance of safety thresholds.

Undoubtedly, prediction of safety thresholds is fraught with various challenges from a scientific, statistical and practical perspective. As shown in Table 1, strengths and weaknesses exist for the different methods currently used for the assessment of safe exposure, whether based on thresholds or not (8). These challenges are often compounded by the restrictive nature of regulatory guidelines for the evaluation of safety pharmacology and toxicity. Typically, experimental protocols for general toxicity used for defining safe exposure ranges in dose escalation (i.e., first-time-in-humans) studies rely on sparse sampling of pharmacokinetic data and other relevant safety measures. Samples are collected according to a pre-defined sampling matrix with a fixed number of animals per time point. Measures of drug exposure are then derived by naive pooling of the data to generate using composite parameters such as AUC and C_{MAX}. Subsequently, these parameters are used to establish the no-adverse-event-level (NOAEL), which determines the maximum allowed exposure during dose escalation in clinical trials (82).

Table 1 Safety thresholds and prediction of risk in humans. Reprinted with permission from Edler *et al.* (7).

	Strengths	Limitations and Weakness
SAR and TTC	• Avoids unnecessary animal testing	• Assumes that structure predicts toxicity • Depends on current exposure estimates for the population
Threshold	• Is simple to apply and readily understood	• Assumes the existence of a threshold • The NOAEL does not exclude biologically significant effects below the sensitivity of the test • The value of the NOAEL depends on experimental conditions such as group size, sensitivity of measurement of the adverse effect, and dose spacing. • Does not make full use of the dose-response information • Uses default UFs
CSAF modelling	• Chemical specific data can be incorporated to reduce uncertainty	• Depends on the validity of the subdivision of the 10-fold factors • Is a data intensive method
Non-threshold	• Linear extrapolation is simple to apply	• Linear extrapolation is thought to be highly conservative. • LMS cannot be validated as a model for low doses and extrapolation is model dependent • Differing balances between reactivity and repair between low and high doses are not accommodated.
BMD	• Makes full use of the dose-response data • Allows confidence limits for point estimates • An optimal experimental design may allow reduction of the number of animals tested (does not require a large number of	• Obtaining consensus defining a benchmark response level for the adverse effect (e.g. 5 or 10%) is difficult • Is not applicable to studies with few dose groups

	animals per group)	
Probabilistic RA	• Uncertainties associated with all aspects of the quantitative methods of the RA process can be taken into account • Appropriate chemical specific information can be incorporated to reduce uncertainty • Provides effect estimates at actual exposure levels	• Requires use of default distributions in most cases
Categorical regression	• Takes all studies into account and not only the most sensitive one • Allows the prediction of a severity effect category at a particular dose (e.g. above ADI)	• Requires toxicological judgement for the categorisation. • The interpretation of fitted model (different endpoints, observer variation etc.) is difficult
PBTK	• Is able to model the time course of the amount of the active compound at the target site • Is possible for any species and for different exposure (e.g. route to route extrapolation) and lifetime conditions • Allows extrapolation from animal to human without having to have human exposure data • Allows target organ dose-response relationships to be used for low-dose extrapolation	• Is a data intensive method • Does not address the dynamics

Given the importance to explore pharmacologically relevant exposure levels in humans, it should be clear that the accuracy of such estimates can become a critical factor during the dose escalation. To date, current guidelines do not describe the implications of variability or bias in these estimates. Yet, the NOAEL is often presented as point estimates to describe the population (22). This ignores variability which can be decomposed into two parts; variability associated with estimation methods and biological variation in pharmacokinetics which arises from inter- and intra-individual differences. Most importantly the exposure estimates from composite measures such as AUC do not allow accurate inferences about the underlying pharmacokinetic processes and individual concentration-effect relationships.

In a previous investigation we have shown that lack of precision exists in exposure estimates derived from the empirical methods currently used for the estimation of toxicokinetic (Sahota et al, unpublished results). One of the main problems is that drug exposure levels observed in satellite animals do not necessarily mirror those assigned to the primary treatment group, in which safety pharmacology and toxicity are evaluated. Evidence form long-standing pharmacokinetic research in pre-clinical species clearly shows that such an approach ignores important differences that may exist between the two experimental groups (11, 12). It is equivalent to assuming that all animals have the same exposure and variability in exposure, i.e., that the underlying physiological processes do not vary between animals. By contrast, the use of a model-based approach enables one to incorporate prior knowledge and additional data from other experiments into the analysis, providing accurate estimates of between- and within-subject variability. This information is essential to ensure a more quantitative, unbiased evaluation of safety pharmacology and toxicology findings.

Arguably, one should not consider only the implications of the statistical method for the analysis and interpretation of safety thresholds, but also question whether experimental protocols are informative enough to allow accurate estimation of the parameters of interest.

In this context, there has been an increase in the awareness about the relevance of optimality concepts for the optimisation and selection of suitable protocol designs for the

evaluation of pharmacokinetic data in conjunction with non-linear mixed-effects modelling. The statistical method was first proposed by Fedorov and later adopted into the PKPD field (83). The approach enables the prospective prediction of parameter precision in the protocol development phase using the expected fisher information matrix (FIM). Variations or adaptations to the original methods have been introduced, which have enable further use of optimality concepts in experimental protocols involving different types of continuous, repeated measurements (84,85). In addition to enhancing the informative value of experimental protocols, the use of optimal design has proven to be an opportunity for reduction in total sample size and consequently in the number of animals required for an experiment (86). Of particular relevance for the evaluation of safety protocols is the possibility of building robust designs to prior uncertainty in pharmacokinetic parameters. Model uncertainty can be explored via sensitivity analysis or by of applying ED-optimality which assumes a prior distribution around the parameters of interest (87).

In the current investigation, simulations are used to illustrate how a model-based approach can be implemented in conjunction with D-optimality software to improve the design of protocols for safety pharmacology and toxicology experiments. It can be anticipated that improved parameter precision and accuracy will allow appropriate dose escalation with less uncertainty about the safety thresholds (20). In fact, our analysis includes an evaluation of the sensitivity to model and parameter uncertainty (21). Furthermore, we also show how to account for the principle of the 3 Rs to ensure that the optimisation procedures do not represent an additional burden to animals required for the experiments (4).

Methods

Currently available software programs have two major limitations for optimising general toxicity protocols. The first is that optimisation is performed with respect to primary model parameters (e.g. CL, Vd). This is restrictive because measures of interest in toxicology are

secondary parameters such AUC and C_{MAX}. For instance, for AUC estimation, the precision of KA is of little importance. Similarly, for most drugs, precise estimation of C_{MAX} will not depend on the precision of CL and peripheral compartment parameters. An optimisation routine that optimises over all parameters may not be suitable either. Ideally, it would be useful to reparameterise the model so that derived measures of exposure are treated as optimisation variables, but this is not always possible as there may be no closed form solution relating primary and secondary model parameters.

The second problem arises from the tendency of software to only provide optimal solutions. In practice there are many other factors to consider (e.g. logistical, ethical, financial, and/or minimal false positive rate) which can be difficult to account for within the optimisation options in a software program. For example, there may be suboptimal designs (in terms of expected parameter precision) that are much more cost effective or ethical. It is therefore important to be able to explore the space of candidate study designs achieving a desired level of precision.

To address the aforementioned problems we proposed to use a simulation-re-estimation approach to study design. However, this is computationally intensive and can quickly become unfeasible when applied to variety of candidate designs and proposal models. For this reason, here we employ a hybrid approach where candidate designs are evaluated in PopED v. 2.10 (University of Uppsala, Sweden) and then expected primary parameter (co)variances are converted to secondary parameter variances using traditional PKPD simulation procedures, as implemented in NONMEM v.6.2 (ICON Development Solutions. Hanover, Maryland).

The studies under consideration were a one week, one month, and three-month general toxicology protocol, in which toxicokinetic data for three different hypothetical drugs were evaluated. Given the pre-defined pharmacokinetic parameters used in the simulations, the true exposure for each individual animal was computed using a variety of measures which were subsequently set as reference for further assessment of the no adverse effect level (NOAEL).

Finally, it should be noted that one of the main issues with the estimation of the NOAEL is that it is limited to the computed exposure at one of the pre-specified experimental doses (22). Consequently, the estimated exposure at any one of the dose levels is a candidate threshold depending on the observed adverse events. To overcome this limitation, the assessment of experimental designs was primarily based on the estimates from secondary parameters (AUC and C_{MAX}) across all treatment groups. In addition, our design space was limited to sampling schedule and number of animals per group to ensure that the NOAEL estimates could be obtained both by NCA and non-linear mixed effects methods. In fact, only experimental designs which allowed for the analysis of the data according to both methods were evaluated.

Given that in typical experimental protocols, three animals are sampled per time point for toxicokinetic analysis, alternative candidate designs were aimed at reducing total sample size, including two or even one animal per sampling time point. These alternative designs represent therefore a reduction in the total number of samples and in the number of animals required per study. Details of the experimental protocols, pharmacokinetic models and optimisation procedures are described in details in the next paragraphs.

Experimental protocols: Three hypothetical drugs were considered to account for differences in disposition properties. We assumed the availability of prior information in the form of single dose pharmacokinetic experiments performed across a range of doses with putative pharmacological activity (1, 3, and 10 mg/kg), in which 8 animals were tested per cohort. The toxicology protocol design was based on an initial set-up commonly used for chronic toxicity evaluation. Four treatment groups (N= 8 per group) receiving oral daily doses of vehicle, 10, 30, and 100 mg/kg/day were tested throughout this set of virtual experiments, which lasted either one week, one month or three months. Satellite groups with 3 animals/time point were used to mimic the dosing conditions in the animals used for the assessment of toxicity (see Figure 1 for a simulation of typical satellite group data). This procedure ensures the availability of more frequent blood samples for toxicokinetics. Blood sampling scheme included four occasions based on feasibility, namely days 1, 8, 25, and 89.

Sampling times on those days were determined by ED-optimality. For the purposes of optimisation, we assumed that all three hypothetical drugs could be fitted by a one-compartment model (model A1) and assumed a 50% CV on all parameters. This was intended to represent standard use of ED-optimality for the optimisation of sampling times. Sampling times were rounded to the nearest 15 minutes.

Pharmacokinetic models: To ensure accurate evaluation of the impact that differences in drug disposition may have on the requirements for experimental design optimisation, three different scenarios were considered in which hypothetical drugs showing on a one-compartment pharmacokinetics with linear and nonlinear (Michaelis-Menten) elimination as well as a two-compartment pharmacokinetics were tested. Parameter values for each scenario are shown in Table 2. In all scenarios, residual variability was assumed to be 15%. Moreover, for the purposes of this exercise, we have assumed a homogeneous population, avoiding the need to explore covariate relationships in any of the models.

Optimisation criteria: See the appendix for background information on the optimality concepts used in this investigation. ED-optimality can be used to incorporate parameter uncertainty into the optimisation process. However, ED optimality only provides an assessment of expected parameter precision and provides no basis for exploration of suboptimal, yet sufficient designs, i.e. reduced designs. Therefore, our decision to use the expected FIM explicitly for the prediction of parameter precision is motivated by a need to have a fast, reliable and flexible method to assess and optimise experimental designs for a model-based analysis whilst adhering to the principle of the 3 Rs. The expected FIM provides a close approximation of expected parameter uncertainty (23,24). In addition, we have favoured the practice of explicitly running the optimisation at different perturbations in model parameters (Table 3). Model parameters were changed in the three PK models tested (one compartment with linear and nonlinear elimination and two compartments), yielding to a total of 27 different models. These models are labelled A1...9, B1....9 and C1....9.

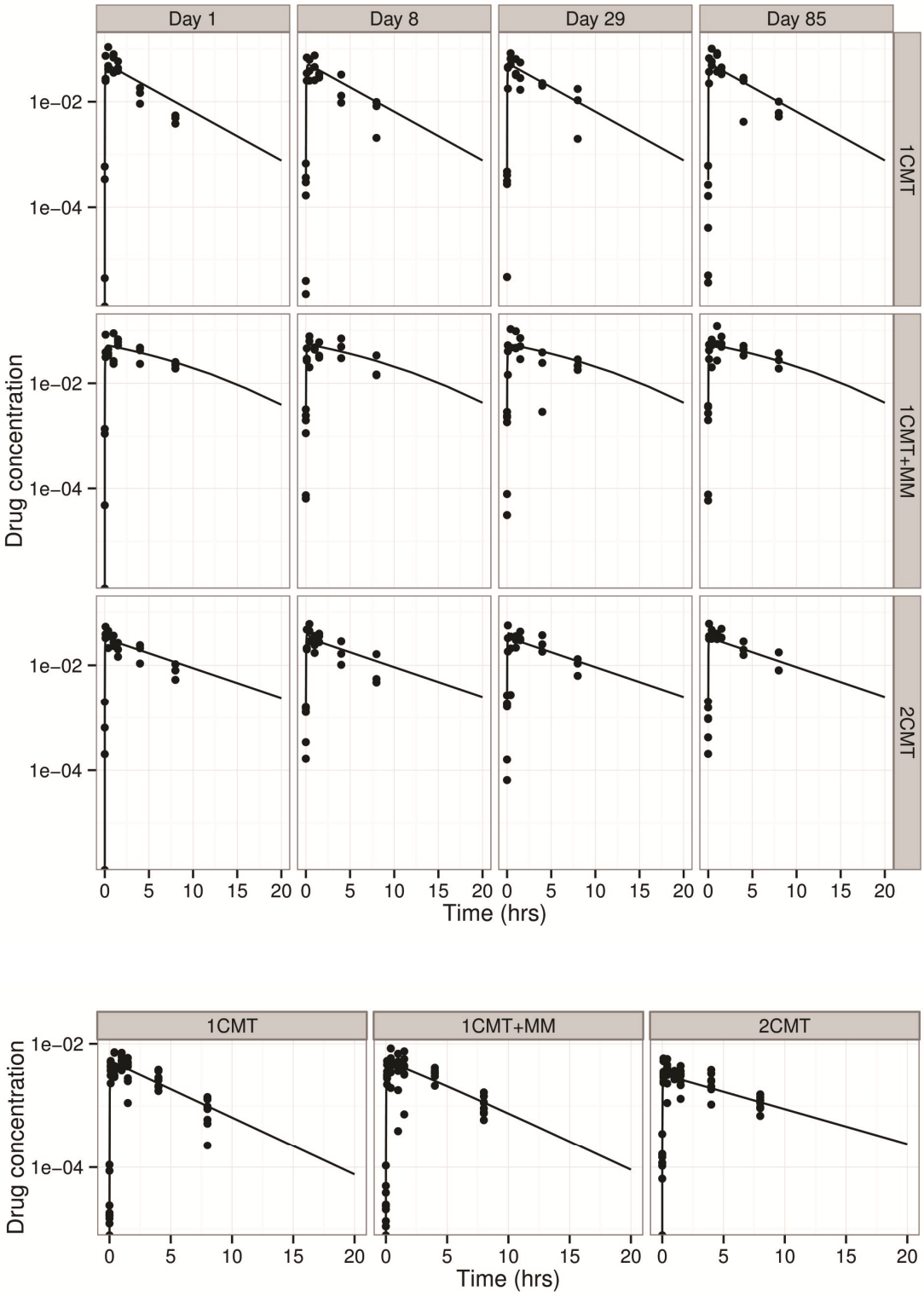

Figure 1: Plots of simulated data for scenarios A1, B1, and C1 overlaid with population prediction (black line). Top panel shows 10mg/kg dosing group using the 3 samples per time point. Bottom panel shows pharmacokinetic profiles at the lower dose level (1 mg/kg) with 8 animals per cohort.

Table 2: Parameters and corresponding between-subject variability used to characterise the pharmacokinetic profiles of hypothetical compounds showing one-compartment, two-compartment and Michaelis-Menten disposition in rats. Doses were defined according to a general toxicology protocol design. Ke: first order rate constant of elimination, Ka: first order rate constant of absorption V: volume of distribution, K_{12}: hybrid constant, K_{21}: hybrid constant; Vmax: maximum metabolic rate ; Km: Michaelis-Menten constant (substrate concentration corresponding to 0.5 V_{max})

MODEL A:

Parameter	Value	BSV (%)
CL (ml/h)	10	20
Ka (h-1)	14.82	50
V (mL)	49	16

MODEL B:

Parameter	Value	BSV (%)
CL (ml/h)	10	20
Ka (h-1)	14.82	50
V (mL)	49	16
K12(h-1)	2.17	16
K21(h-1)	3.554	69

MODEL C:

Parameter	Value	BSV (%)
Vmax (mg/h)	0.3	20
Ka (h-1)	14.82	50
V (mL)	49	16
Km(mg/L)	30	0 FIX

Table 3: Perturbations in the parameters for the three different pharmacokinetic models. CL: clearance, Ka: first order rate constant of absorption V: volume of distribution, Vmax: maximum metabolic rate.

Model	KA	V	CL	Model	KA	V	CL	Model	KA	V	VMAX
A1	-	-	-	B1	-	-	-	C1	-	-	-
A2	-	+50%	+50%	B2	-	+50%	+50%	C2	-	+50%	+50%
A3	-	+50%	-50%	B3	-	+50%	-50%	C3	-	+50%	-50%
A4	-	-50%	+50%	B4	-	-50%	+50%	C4	-	-50%	+50%
A5	-	-50%	-50%	B5	-	-50%	-50%	C5	-	-50%	-50%
A6	-80%	+50%	+50%	B6	-80%	+50%	+50%	C6	-80%	+50%	+50%
A7	-80%	+50%	-50%	B7	-80%	+50%	-50%	C7	-80%	+50%	-50%
A8	-80%	-50%	+50%	B8	-80%	-50%	+50%	C8	-80%	-50%	+50%
A9	-80%	-50%	-50%	B9	-80%	-50%	-50%	C9	-80%	-50%	-50%

All evaluations were performed in PopED v.2.10 (University of Uppsala, Sweden) (88), a software developed in O-Matrix® (Harmonic Software Inc., Seattle, WA, USA). Data manipulation and statistical and graphical summaries were performed in R 2.10.0 (26). In our analysis, the expected FIM was used to compute the expected covariance matrix from which, the expected precision of primary pharmacokinetic parameters was quantified (89,90).

The expected precision of the derived parameters of interest, namely AUC and C_{MAX}, were calculated from the expected covariance matrix of primary parameters in NONMEM 6.2 (ICON Development Solutions. Hanover, Maryland) (27). First, 1000 pharmacokinetic profiles were simulated from the primary parameters uncertainty distributions by including the covariance information in the $PRIOR subroutine. For each pharmacokinetic profile, the AUC and C_{MAX} were calculated as follows:

$$AUC = \int_{t-24}^{t} C_p dt$$

$$C_{MAX} = \max\left(\{C_p(s): t-24 < s < t\}\right)$$

where individual predicted drug concentrations are denoted by $C_p(t)$.

The expected precision (standard error) of the parameters was then summarised. Adequate precision was defined as expected CV% < 40%. Absolute changes in expected precision of less than 10% were deemed biologically irrelevant.

Results

Our analysis shows that optimal design concepts can be used in toxicology research to improve the precision of the parameters of interest whilst allowing for a reduction in the total number of animals required per experiment. As shown in figure 1, plots of the

simulated profiles for a typical individual together with simulated samples, representing "observed" data are depicted to illustrate the impact of different disposition characteristics on the concentration vs. time profiles.

The optimised sampling times for all scenarios were 0.25, 0.5,0.75, 1, 1.5, 2, 8 and 24 hours after dosing. Results show that for all designs the precision of AUC and C_{MAX} associated with a reduced sample size of 2/3 from the initial sample size resulted in an acceptable loss of precision (the absolute difference in expected precision was <10% for all scenarios for sample size reduction of 2/3). Therefore, optimised protocols result in a reduction of up to 2/3 in the number of animals utilised in toxicokinetic experiments.

An overview of the point estimates and coefficient of variation (CV%) obtained for AUC and C_{MAX} is presented in Table 5. The differences in parameter precision associated with varying sample size, including the NOAEL, is summarised for each model in Figures 2, 3 and 4). We show how precision changes when one or two animals are sampled at each time point instead of using 3 animals per sampling time point. Interestingly, the expected precision was very high for the one-compartmental model but there was less precision for the two-compartmental model, where a distribution phase is evident. In addition, our analysis reveals that metabolic saturation, as described by Michaelis-Menten kinetics does not further affect the precision of parameter estimates. Further assessment of the precision of the primary parameters indicates that the parameters governing peripheral compartment distribution will be the least precisely estimated, with a loss of precision as high as 75% for some parameter perturbations. Between-subject variability was also found to be imprecisely estimated and would have to be fixed to 0 for some parameters during data analysis. Yet, despite these differences, AUC and C_{MAX} imprecision was <36% for the two-compartmental models.

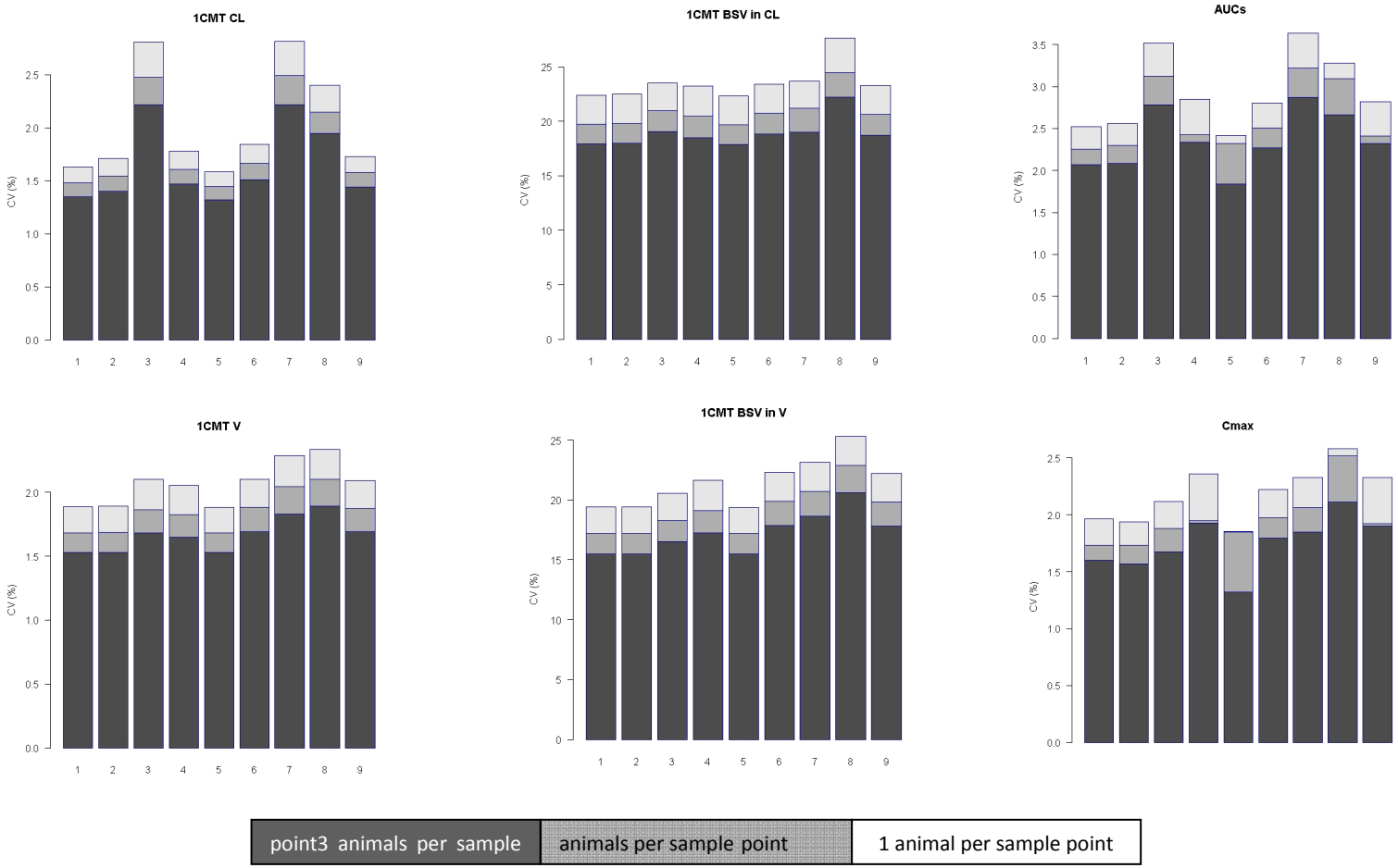

Figure 2. Bar charts of CVs of selected parameters for models A.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

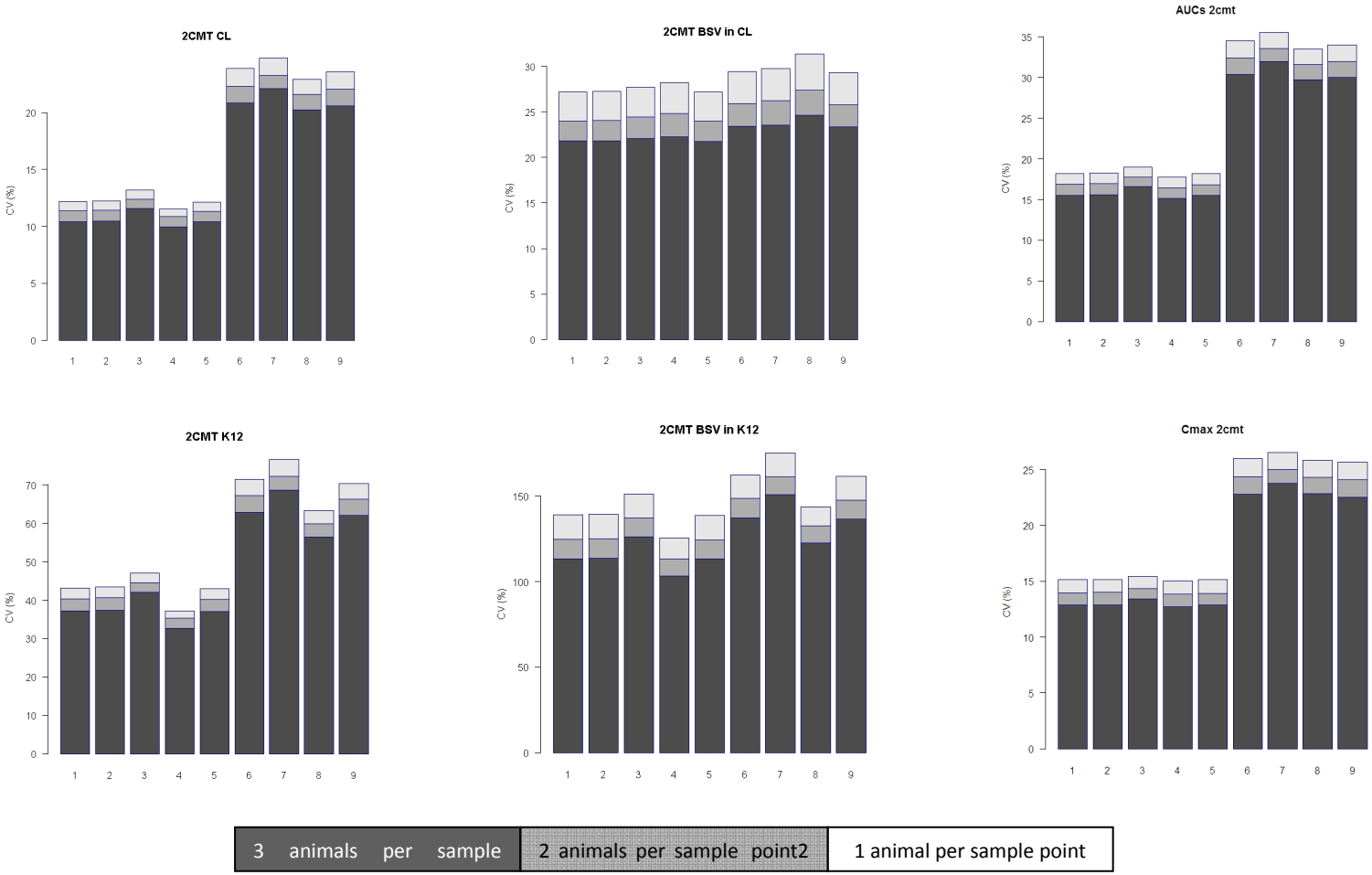

Figure 3. Bar charts of CVs of selected parameters for models B.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

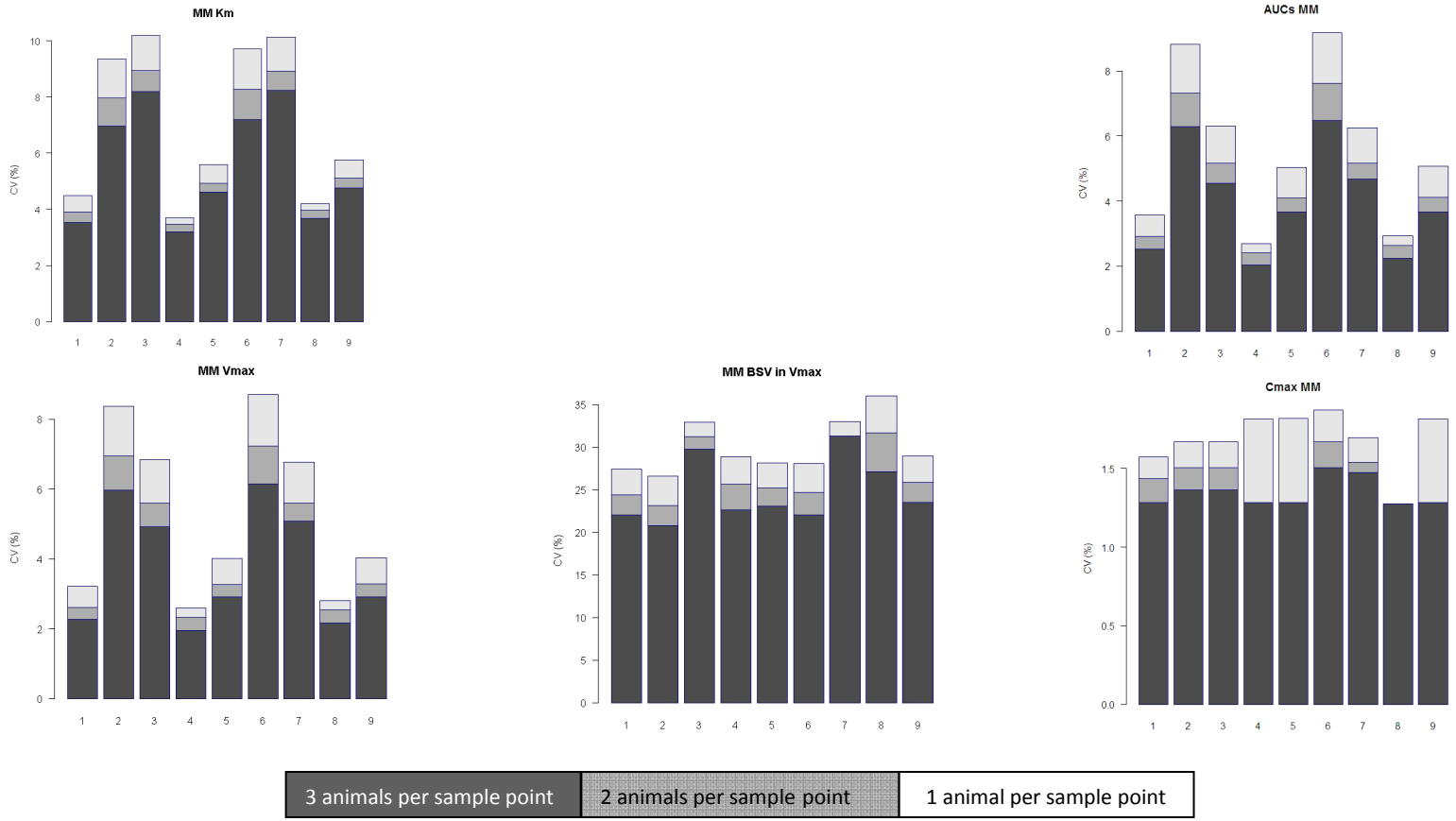

Figure 4. Bar charts of CVs of selected parameters for models C.x, where x range from 1-9 and is indicated on the x-axis. The y-axis shows expected precision of the various scenarios.

Discussion

Experimental protocols based on repeated-dose treatment arms are essential for accurate inferences about the risk associated with the exposure to new chemical entities in the early phase of clinical development. These studies provide the basis for the calculation of safety thresholds such as the no-observed-adverse-effect level (NOAEL) or lowest-observed-adverse-effect level (LOAEL), which are used to extrapolate the concentration or exposure above which adverse effects can be expected in humans (82,91).

Despite the efforts and attention given to different methodologies for the estimation of such safety thresholds, it is now acknowledged that the use of NOAEL or LOAEL as traditional thresholds or point of departure for risk assessment has significant limitations. The NOAEL and LOAEL are determined by the selected dose levels and intervals used in an experimental protocol.

To date, these measures remain a requirement for regulatory purposes (2). However, there is a wide consensus that they do not mathematically relate to the underlying exposure-response curve (92). In addition, it has been shown that differences in protocol design can influence the precision and accuracy of the parameters of interest, yielding biased NOAEL and LOAEL estimates. In fact, the bench mark dose (BMD) as the threshold or point of departure has been proposed as an alternative method to avoid many of these pitfalls (41). Unfortunately, similar challenges exist with regard to the accuracy and precision of estimates obtained by the BMD (18,93). The experimental data are not integrated nor parameterised in a mechanistic manner so as to benefit from the advantages of a model-based approach.

Whilst risk assessment methods need undoubtedly to incorporate mechanistic aspects of drug action to ensure better characterisation of potential hazards to humans, it should be noted that improvements are also required from a statistical perspective. Thus far empiricism and regulatory-related issues have dominated traditional toxicological testing

paradigms (32-35). Minimal efforts have been made to introduce optimality concepts in experimental design as a means to increase accuracy and precision of the parameters of interest.

In this investigation we have attempted to show the feasibility of implementing a model-based approach in conjunction with optimal design based on techniques, which have been developed for the field of pharmacokinetics for more than two decades ago (13,36,37). By considering a number of hypothetical scenarios in which drugs with different disposition properties were simulated, we have demonstrated that accurate estimates of AUC and C_{MAX} can be obtained for drugs showing different pharmacokinetic profiles. Our results also highlight the impact of optimisation procedures on the estimation of secondary parameters. We have shown that even when precision of the primary pharmacokinetic parameter is poor, as in the case of parameters governing distribution into peripheral compartments, the precision of the secondary parameters remains unaffected. This can be attributed to the fact that the selected candidate designs systematically yield estimates of clearance and volume of distribution with acceptable precision. These two parameters ultimately determine systemic exposure and peak concentrations, respectively.

Although it may seem a disadvantage to use model-dependent estimates for the assessment of safety thresholds, this approach presents various important advantages (38-39). First, it is unbiased and predictive, allowing for the incorporation of the physiological factors underlying the pharmacokinetic properties of the drug under investigation. Moreover, it enables ne to integrate prior information, including data from other experiments. We anticipate that many areas in toxicology research which can benefit from such an approach. New methodology does not necessarily mean that human safety will be placed at risk. On the contrary, newer methods provide an opportunity to remove much of the guess work involved with older methodologies, which rely on assumptions which clearly prevent the uptake of evolving knowledge about pharmacokinetic and pharmacodynamic properties of a drug.

Methodological aspects

In assessing and optimising the protocol we found that existing routines in optimality software were insufficient to meet our assessment criteria. In particular, existing software did not enable the assessment and optimisation over arbitrary secondary parameters, and did not allow for the impact of parameter perturbations on expected precision to be assessed. The alternative brute force approach to account for these limitations would have been to perform multiple simulation-re-estimation procedures across our design and model space. However, this would have involved extensive computation times. Our approach instead consisted of FIM evaluations followed by calculation of the expected secondary parameter precision. This exercise ultimately showed that optimisation can be performed on secondary parameters of interest, and minimally sufficient designs can be obtained. Both of these procedures are computationally inexpensive. Our approach therefore enables exploration of large design and model spaces without the aforementioned limitations in current optimality software.

Limitations

Our work does involve a number of assumptions, which may represent potential theoretical and practical limitations. First, it should be noted that we have constrained ourselves to candidate designs that enable estimation of exposure using non-compartmental methods for each treatment group. Further gains in terms of reduced burden and/or parameter precision are likely to be achieved if a model-based analysis was the only intended analysis of the data.

Another requirement is the availability of a well-defined population pharmacokinetic model, which is feasible, but in practice not used in routine pre-clinical research. It should be clear that the computation of expected (co)variance by means of the FIM, cannot directly account for the possibility of unidentifiably of parameters. Hence, the validity of any optimisation procedures implies accurate knowledge of the pharmacokinetic properties and corresponding parameterisation. Parameter unidentifiability will likely manifest in terms of

large standard errors, high correlations in the correlations and/or large differences in eigenvalues. On the other hand, optimal design does tackle another common issue observed during data fitting and parameter estimation, i.e., numerical unidentifiability, which may be caused by poor experimental design.

An additional assumption is that parameter estimates will be unbiased. This assumption may not hold true for more complex models, but the reader should be aware that this issue may be equally important when non-compartmental methods are used to describe complex pharmacokinetic profiles, as for instance in the case of metabolic inhibition or drugs with long elimination half-life (40). To ensure further characterisation of bias, a full bootstrap (simulation-re-estimation) procedure is recommended. Lastly, one should realise the implications of our own objectives, i.e., to compare designs which are suitable for both non-compartmental and model-based methods. Further gains in terms of reduced burden and/or parameter precision are likely to be achieved if a model-based analysis was the only intended analysis of the data.

In summary, it can be concluded that despite the biological debate about the relevance of safety thresholds, the accuracy and precision of estimates are essential to ensure appropriate interpretation of experimental findings and make inferences about risk in humans. We have shown that the use of a model-based approach is critical for appropriate data integration and informative value of experimental protocols. Our work also demonstrates that population size is not the critical variable when evaluating precision and accuracy of the parameters of interest. This feature allows for comparable results to be obtained with considerable lower number of animals and consequently reduction in the cost of experiments. Overall, these results make the need to explore the requirements for further implementation of optimal design in toxicology research an ethical and scientific imperative.

Reference List

1. ICH Expert Working Group Guideline S3A. Note for guidance on toxicokinetics: The assessment of systemic exposure in toxicity studies, 1994.

2. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry M3(R2). Nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. 2010.

3 Hartford CG, Petchel KS, Mickail H, Perez-Gutthann S, McHale M, Grana JM, Marquez P. Pharmacovigilance during the pre-approval phases: an evolving pharmaceutical industry model in response to ICH E2E, CIOMS VI, FDA and EMEA/CHMP risk-management guidelines. Drug Saf. 2006; 29(8): 657-73.

4. Cooper AJ, Lettis S, Chapman CL, Evans SJ, Waller PC, Shakir S, Payvandi N, Murray AB. Developing tools for the safety specification in risk management plans: lessons learned from a pilot project. Pharmacoepidemiol Drug Saf. 2008; 17(5):445-54. doi: 10.1002/pds.1576.

5. Aronson JK, Ferner RE. Joining the DoTS: new approach to classifying adverse drug reactions. BMJ 2003; 327:1222-1225.

6. DeGeorge JJ. Food and Drug Administration viewpoints on toxicokinetics: the view from review. Toxicol Pathol. 1995; 23(2): 220-5.

7. Bessems JG, Geraets L. Proper knowledge on toxicokinetics improves human hazard testing and subsequent health risk characterisation. A case study approach. Regul Toxicol Pharmacol. 2013; 67(3):325-34.

8. Edler L, Poirier K, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K, Wurtzen G. Mathematical modelling and quantitative methods. Food Chem Toxicol. 2002; 40:283-326.

9. Contrera J F, Matthews EJ, Kruhlak NL, Benz RD. Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul Toxicol Pharmacol. 2004; 40:185-206.

10. Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol. 2005; 42:265-274.

11. Singh SS. Preclinical pharmacokinetics: an approach towards safer and efficacious drugs. Curr Drug Metab. 2006; 7(2):165-82.

12. Ding X, He M, Kulkarni R, Patel N, Zhang X. Investigation of clinical pharmacokinetic variability of an opioid antagonist through physiologically based absorption modeling. J Pharm Sci. 2013; 102(8):2859-74.

13. Fedorov V. Optimal experimental design. WIREs Comp Stat 2010; 2: 581-589.

14. Retout S, Mentre F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Stat Med. 2002; 21:2623-2639.

15. Duffull SB, Mentre F, Aarons L. Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm Res. 2001; 18:83-89.

16. Waterhouse TH, Redmann S, Duffull SB, Eccleston JA. Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. J Pharmacokinet Pharmacodyn. 2005; 32:521-545.

17. Ette EI, Howie CA, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in preclinical pharmacokinetic studies. Pharm Res. 1995; 12:729-737.

18. Tod M, Rocchisani JM. Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics. J Pharmacokinet Biopharm. 1997; 25:515-537

19. Hooker AC, Foracchia M, Dodds MG, Vicini P. An evaluation of population D-optimal designs via pharmacokinetic simulations. Ann Biomed Eng 2003; 31:98-111.

20. Balls M, Goldberg AM, Fentem JH, Broadhead CL, Burch RL, Festing MFW, Frazier JM, Hendriksen CFM, Jennings M, van der Kamp MDO, Morton DB, Rowan AN, Russell C, Russell WMS, Spielmann H, Stephens ML, Stokes WS, Straughan DW, Yager JD, Zurlo J , van Zutphen BFM. The Three Rs: the way forward. Alternatives to Laboratory Animals 1995; 23: 838-866.

21. Nestorov I. Modelling and simulation of variability and uncertainty in toxicokinetics and pharmacokinetics. Toxicol Lett. 2001; 120:411-420.

22. Ogungbenro K, Gueorguieva I, Majid O, Graham G, Aarons L. Optimal design for multiresponse pharmacokinetic-pharmacodynamic models - dealing with unbalanced designs. J Pharmacokinet Pharmacodyn. 2007; 34:313-331.

23. Thai H, Mentre F, Holford NH, Veyrat-Follet C, Comets E. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics . J Pharmacokinet Pharmacodyn. (2014) 41:15–33

24. Hooker A, Vicini P. Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. AAPS J. 2005; 7:E759-E785.

25. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. Comput Methods Programs Biomed. 2004; 74:29-46.

26. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

27. Beal, S., Sheiner, L.B., Boeckmann, A., & Bauer, R.J., NONMEM User's Guides. (1989-2009), Icon Development Solutions, Ellicott City, MD, USA, 2009.

28. Huang C, Zheng M, Yang Z, Rodrigues AD, Marathe P. Projection of exposure and efficacious dose prior to first-in-human studies: how successful have we been? Pharm Res. 2008; 25:713-726.

29. Gibbs JP. Prediction of exposure-response relationships to support first-in-human study design. AAPSJ. 2010; 12:750-758.

30. Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C, Spitzer HL. Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Regul Toxicol Pharmacol. 1995; 21:296-306.

31. Budtz-Jorgensen E, Grandjean P, Keiding N, White RF, Weihe P. Benchmark dose calculations of methylmercury-associated neurobehavioural deficits. Toxicol Lett. 2000; 112-113: 193-199.

32. indgren S, Bass AS, Briscoe R, Bruse K, Friedrichs GS, Kallman MJ, Markgraf C, Patmore L, Pugsley MK. Benchmarking safety pharmacology regulatory packages and best practice. J Pharmacol Toxicol Methods. 2008; 58(2):99-109.

33. Rowbotham AL, Gibson RM. Exposure-driven risk assessment: applying exposure-based waiving of toxicity tests under REACH. Food Chem Toxicol. 2011; 49(8):1661-73.

34. Kooijman M, van Meer PJ, Moors EH, Schellekens H. Thirty years of preclinical safety evaluation of biopharmaceuticals: Did scientific progress lead to appropriate regulatory guidance? Expert Opin Drug Saf. 2012; 11(5):797-801

35. Schellekens H. Overregulation and unnecessary animal testing: requirements for market approval of biopharmaceuticals too rigid. Ned Tijdschr Geneeskd. 2012; 156(50):A5723.

36. D'Argenio, DZ. Optimal sampling times for pharmacokinetic experiments. J Pharmacokinet Biopharm. 1981; 9:739-756

37. al Banna MK, Kelman AW, Whiting B. Experimental design and efficient parameter estimation in population pharmacokinetics. J Pharmacokinet Biopharm. 1990; .18:347-360.

38. Beatty DA, Piegorsch WW. Optimal statistical design for toxicokinetic studies. Stat Methods Med Res. 1997; 6(4):359-76.

39. Dodds MG, Hooker AC, Vicini P. Robust population pharmacokinetic experiment design. J Pharmacokinet Pharmacodyn. 2005; 32(1):33-64.

40. Lehr T, Staab A, Trommeshauser D, Schaefer HG, Kloft C. Semi-mechanistic population pharmacokinetic drug-drug interaction modelling of a long half-life substrate and itraconazole. Clin Pharmacokinet. 2010; 49(1):53-66.

41. Nyberg J, Karlsson MO, Hooker AC. Simultaneous optimal experimental design on dose and sample times. J Pharmacokinet Pharmacodyn. 2009; 36:125-145.

42. Wang J, Endrenyi L. A computationally efficient approach for the design of population pharmacokinetic studies. J Pharmacokinet Biopharm. 1992; 20:279-294.

Appendix

In an optimal design exercise, design variables are variables that describe properties of the biological system, drug or experimental protocol which can be changed to explore their impact on the information contents of the experiment. Typically these include dose, sampling scheme, number of samples, number of individuals or other covariates (94). Even though the number of animals is constrained (88), the main use of this technique is to optimise sampling times. It has been shown that sample times can have significant influence in the accuracy and precision of parameters (95,96). By optimising sampling times it is possible therefore to improve the overall efficiency of PK experiments (96,97).

Here we summarise the statistical framework for the evaluation and optimisation of experimental designs using D-optimality. There are various software programs for optimal design, making them equally suitable for the purposes of this type of analysis. They differ primarily in the features available for optimisation and in the optimisation method.

Statistical summary

There are various numerical methods to fit a model to data. The mostly commonly used is the maximum likelihood (ML) estimator. The maximum likelihood is calculated by maximising the following likelihood function (L):

$$L(\theta) = p(D|\theta)$$

where θ is the vector of parameters, *D* is the data. The results of a maximum likelihood estimation are $\hat{\theta}$, the maximum likelihood estimate and cov($\hat{\theta}$), the covariance matrix determining the parameter precision. The information contents within the study data, *D* is what determines cov($\hat{\theta}$). Prior to running the experiment, assuming the availability of a model, it is possible to compute an expected covariance matrix by the use of the Cramer-Rao inequality:

$$cov(\hat{\theta}) \geq \frac{1}{FIM(\hat{\theta})}$$

where the Fisher Information Matrix (FIM) is given by

$$FIM(\hat{\theta}) = E\left[\left(\frac{\partial}{\partial\theta}L(\theta)\right)^{T}\left(\frac{\partial}{\partial\theta}L(\theta)\right)\right]$$

Although this function constrains the lower bound of cov($\hat{\theta}$), in practice such a lower bound is reached as indicated by comparisons with bootstrapped expected covariance estimates (98,99). Thus, by computing the FIM of a given design, under the assumption of no or minor model and parameter misspecification, one can estimate the covariance matrix and consequently assess parameter precision values. By maximising the determinant of the FIM over design variables, such, as for instance the sampling schedule, it is possible to identify experimental conditions or design(s) that maximise the expected parameter precision.